PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-104791
(43)Date of publication of application : 21.04.1995

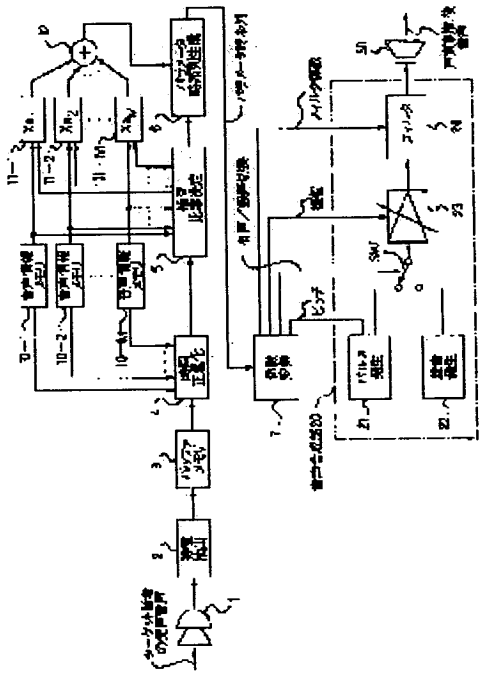
(51)Int.Cl. G10L 7/02

(21)Application number : 05-247828
(22)Date of filing : 04.10.1993

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK
(72)Inventor : IWAHASHI NAOTO
KOSAKA YOSHINORI

(54) VOICE QUALITY CONTROL TYPE VOICE SYNTHESIZING DEVICE

(57)Abstract:
PURPOSE: To obtain a device which can synthesize voice after voice quality control to voice inputted with a higher degree of freedom through a simple constitution by performing an interpolating process for data on voice spectra of plural speakers by using a specific interpolation rate and outputting data on a voice spectrum having specific voice quality, and synthesizing the voice on the basis of the data.
CONSTITUTION: Voice information memories 10-1-10-M store the data of the voice spectra of plural M speakers in advance respectively, the voice of a target speaker is inputted, and the voice spectra of the M speakers are interpolated and mixed so as to approximate the input voice, thereby vocalizing a voice close to the voice quality of the voice of the target speaker. And, a voice having specific different quality is vocalized by varying the interpolation ratio. Here, a cepstrum coefficient, etc., having relatively excellent interpolation characteristics is used as a spectrum parameter for interpolation, and the interpolation ratio of the parameter is varied to synthesize and output a voice having different quality.



LEGAL STATUS

[Date of request for examination]	04.10.1993
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	2951514
[Date of registration]	09.07.1999
[Number of appeal against examiner's decision	

of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right] 09.07.2003

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（A）**　(11)特許出願公開番号

特開平7－104791

(43)公開日　平成７年(1995)４月21日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ　7/02		D		

審査請求　有　　請求項の数４　ＯＬ　（全　６　頁）

(21)出願番号　特願平5－247828

(22)出願日　平成５年(1993)10月４日

(71)出願人　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地

(72)発明者　岩橋　直人
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(72)発明者　匂坂　芳典
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

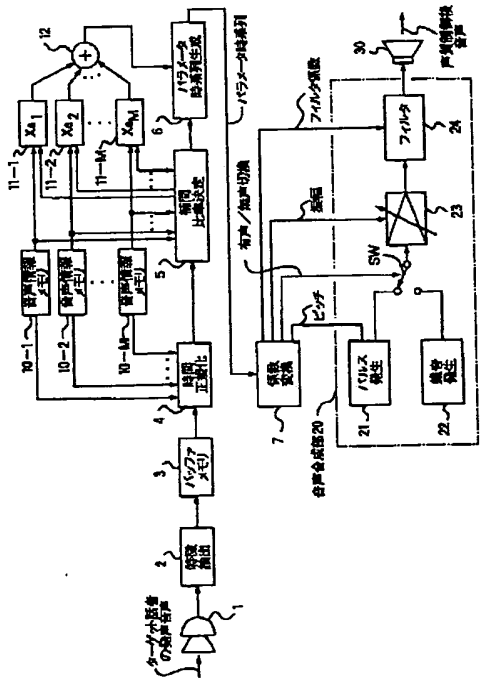
(74)代理人　弁理士　青山　葆　（外２名）

(54)【発明の名称】　声質制御型音声合成装置

(57)【要約】

【目的】　従来例に比較して非常に簡単な構成を有し、しかもより高い自由度を有して入力された音声に対して声質制御した後の音声を合成することができる音声合成装置を提供する。

【構成】　複数の話者の音声スペクトルのデータを予め記憶する記憶装置と、記憶装置から複数の話者の音声スペクトルのデータを読み出し、所定の補間比率を用いて上記複数の話者の音声スペクトルのデータに対して内挿処理を実行して所定の声質を有する音声スペクトルのデータを出力する処理回路と、処理回路から出力される音声スペクトルのデータに基づいて音声を合成して出力する音声合成回路とを備える。さらに、ターゲット話者の音声に基づいて音声スペクトルのデータを抽出し、抽出された音声スペクトルのデータが処理回路から出力される音声スペクトルのデータに近似するように補間比率を演算して設定する演算回路とを備える。



1

【特許請求の範囲】

【請求項1】　複数の話者の音声スペクトルのデータを予め記憶する記憶手段と、

上記記憶手段から複数の話者の音声スペクトルのデータを読み出し、所定の補間比率を用いて上記複数の話者の音声スペクトルのデータに対して内挿処理を実行して所定の声質を有する音声スペクトルのデータを出力する処理手段と、

上記処理手段から出力される音声スペクトルのデータに基づいて音声を合成して出力する音声合成手段とを備えたことを特徴とする声質制御型音声合成装置。

【請求項2】　声質制御型音声合成装置はさらに、

ターゲット話者の音声に基づいて音声スペクトルのデータを抽出する特徴抽出手段と、

上記特徴抽出手段によって抽出された音声スペクトルのデータが上記処理手段から出力される音声スペクトルのデータに近似するように上記補間比率を演算して上記処理手段に設定する演算手段とを備えたことを特徴とする請求項1記載の声質制御型音声合成装置。

【請求項3】　声質制御型音声合成装置はさらに、

上記補間比率を入力して上記処理手段に設定する入力手段を備えたことを特徴とする請求項1記載の声質制御型音声合成装置。

【請求項4】　上記音声スペクトルのデータは、ケプストラム係数又は対数面積比であることを特徴とする請求項1、2又は3記載の声質制御型音声合成装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、声質を制御して音声合成することができる声質制御型音声合成装置に関する。

【0002】

【従来の技術及び発明が解決しようとする課題】男性の声から女性の声に又は女性の声から男性の声に変換する方法（以下、第1の従来例という。）が、例えば、箱田和雄，“極制御による男女声変換法の検討”，日本音響学会講演論文集，2－6－13，pp213－214，昭和62年10月に開示されている。この第1の従来例においては、ＬＰＣ分析で得られる極周波数を用いて、母音情報の代わりに極周波数値を用いて極の変換を行うとともに、変換に伴うスペクトルの傾きの変動を2次のフィルタを用いて吸収する方法を提案している。しかしながら、男性の声から女性の声に又は女性の声から男性の声への変換のみで、例えば男性から男性への変換で、異なる個人性を有する音声を発生することはできない。

【0003】また、アナウンサーなどの明瞭は音声の物理的特性の1つとして、ホルマント周波数のダイナミックな変動に着目し、特に声の明瞭性を改善するために、ホルマント周波数の時間変化を制御する方法（以下、第2の従来例という。）が、例えば、都木徹，桑原尚夫，“ホルマント変化の強調・抑圧による声質制御”，日本

2

音響学会講演論文集，1－4－12，pp145－146，昭和61年10月に開示されている。この第2の従来例は、次のステップを有する。

（a）所定の標本化周波数でＤ／Ａ変換した音声信号を、所定のフレーム幅、フレーム周期、及び男女で異なる分析次数で線形予測分析し、予測係数との残差を計算する。

（b）各フレーム毎に予測係数からホルマント周波数を算出し、従来の粕谷の方法（“線形予測分析法で得られる極周波数からのホルマント周波数選択アルゴリズム”，電子通信学会論文誌，Ｖｏｌ．Ｊ66－Ａ，Ｎｏ．11，ｐｐ．1144－1145，1983年11月参照。）により、母音部の第1乃至第3ホルマントの軌跡を求める。

（c）得られたホルマント軌跡に対して、ある音声のホルマント周波数の時間変化を示す所定の式に適用し、各フレーム毎に新たなホルマント軌跡を求め、その値から合成に用いる予測係数を算出する。なお、ここで、第4以上のホルマント及び無声音部、有声子音部は変更しない。

（d）新たな予測係数と最初に求めた残差から合成音を発生する。

この第2の従来例においては、ホルマント周波数のダイナミックスを強調又は抑制するのみなので、声の明瞭性を改善することはできるが、第1の従来例と同様に、異なる個人性を有する音声を発生することはできず、声質制御の自由度が小さいという問題点があった。

【0004】さらに、ある特定話者からターゲット話者への声質の変換方法（以下、第3の従来例という。）が、M.Abe et al.，“Voice Conversion through vector quantization”,Proc. ICASSP'88,pp.655-658,1988年に開示されている。この第3の従来例においては、いわゆるコードベクトル・マッピング手法に基づいて特徴パラメータのベクトル量を制御し、これらのマッピングは音声スペクトルに対する適切な拘束なしに学習データから計算されていたために、ターゲット話者への適切なマッピング関数を求めるときに、ターゲット話者による大量の発声データを必要とし、極めて大きな記憶装置を設ける必要があるという問題点があった。

【0005】本発明の目的は以上の問題点を解決し、従来例に比較して非常に簡単な構成を有し、しかもより高い自由度を有して声質制御して音声合成することができる音声合成装置を提供することにある。

【0006】

【課題を解決するための手段】本発明に係る請求項1記載の声質制御型音声合成装置は、複数の話者の音声スペクトルのデータを予め記憶する記憶手段と、上記記憶手段から複数の話者の音声スペクトルのデータを読み出し、所定の補間比率を用いて上記複数の話者の音声スペクトルのデータに対して内挿処理を実行して所定の声質

を有する音声スペクトルのデータを出力する処理手段と、上記処理手段から出力される音声スペクトルのデータに基づいて音声を合成して出力する音声合成手段とを備えたことを特徴とする。

【０００７】また、請求項２記載の声質制御型音声合成装置は、請求項１記載の声質制御型音声合成装置において、さらに、ターゲット話者の音声に基づいて音声スペクトルのデータを抽出する特徴抽出手段と、上記特徴抽出手段によって抽出された音声スペクトルのデータが上記処理手段から出力される音声スペクトルのデータに近似するように上記補間比率を演算して上記処理手段に設定する演算手段とを備えたことを特徴とする。

【０００８】さらに、請求項３記載の声質制御型音声合成装置は、請求項１記載の声質制御型音声合成装置において、さらに、上記補間比率を入力して上記処理手段に設定する入力手段を備えたことを特徴とする。

【０００９】またさらに、請求項４記載の声質制御型音声合成装置は、請求項１、２又は３記載の声質制御型音声合成装置において、上記音声スペクトルのデータは、ケプストラム係数又は対数面積比であることを特徴とする。

【００１０】

【作用】音声スペクトルを変更することにより、声質を制御するためには、音声のスペクトル構造及びそのダイナミックスに存在するある種の特徴を適切にモデル化し、そのモデル化に基づいたスペクトルの制御を行うことが望ましいと考えられる。しかしながら、音声スペクトルのモデル化を直接に、声門運動や声道形状に基づいて行った音声合成方式では、ホルマント合成で人手による精密な制御を行った数例を除き、自動的な手段を用いて高品質な合成音を発声させることはできていない。そこで、本発明者は、このような物理的モデルを直接的に用いる代わりに複数の人数の音声スペクトルそのものをノン・パラメトリックな音声スペクトルモデルとみなし、これをスペクトル制御の拘束条件として用いることを考えた。新しい音声スペクトルは、以下に詳細後述するように、複数の人数の音声スペクトルを線形に内挿することで求める。

【００１１】請求項１記載の声質制御型音声合成装置においては、上記処理手段は、上記記憶手段から複数の話者の音声スペクトルのデータを読み出し、所定の補間比率を用いて上記複数の話者の音声スペクトルのデータに対して内挿処理を実行して所定の声質を有する音声スペクトルのデータを出力し、次いで、上記音声合成手段は、上記処理手段から出力される音声スペクトルのデータに基づいて音声を合成して出力する。

【００１２】また、請求項２記載の声質制御型音声合成装置においては、請求項１記載の声質制御型音声合成装置において、さらに、上記特徴抽出手段は、ターゲット話者の音声に基づいて音声スペクトルのデータを抽出

し、上記演算手段は、上記特徴抽出手段によって抽出された音声スペクトルのデータが上記処理手段から出力される音声スペクトルのデータに近似するように上記補間比率を演算して上記処理手段に設定する。これによって、上記ターゲット話者の音声に近似した音声を上記音声合成手段によって合成することができる。

【００１３】さらに、請求項３記載の声質制御型音声合成装置においては、請求項１記載の声質制御型音声合成装置において、さらに、上記入力手段は、上記補間比率を入力して上記処理手段に設定する。従って、上記補間比率を変更して種々の声質を有する音声を上記音声合成手段によって発声させることができる。

【００１４】またさらに、請求項４記載の声質制御型音声合成装置においては、請求項１、２又は３記載の声質制御型音声合成装置において、上記音声スペクトルのデータは、好ましくは、ケプストラム係数又は対数面積比である。

【００１５】

【実施例】以下、図面を参照して本発明に係る実施例について説明する。図１は本発明に係る一実施例である声質制御型音声合成装置のブロック図である。

【００１６】本実施例の音声合成装置は、複数の人数の音声スペクトルを線形に内挿する処理のために、予め音声情報メモリ１０－１乃至１０－Mにそれぞれ複数M人の話者の音声スペクトルのデータを格納し、ターゲット話者の音声を入力してその音声に近似するように、上記複数M人の話者の音声スペクトルを内挿混合することにより、上記ターゲット話者の音声の声質に近い音声を発声させる一方、上記内挿比率を変更することによって所定の異なる声質を有する音声を発声させることを特徴としている。ここで、内挿するためのスペクトル・パラメータとして、比較的良好な補間特性を有するケプストラム係数又は対数面積比を用い、パラメータの内挿比率を変更することにより、異なる声質を有する音声を合成して出力させる。

【００１７】ターゲット話者の発声音声はマイクロホン１に入力されて音声信号に変換された後、特徴抽出部２に入力される。一方、音声情報メモリ１０－１乃至１０－Mにそれぞれ、複数M人の話者の音声スペクトルのデータを格納されている。ここで、音声スペクトルのデータは、音声スペクトルの振幅の時系列データ及び例えば１６次のケプストラム係数の時系列データを含む。

【００１８】特徴抽出部２は、入力された音声信号をＡ／Ｄ変換した後、例えばＬＰＣ分析を実行し、対数パワー、１６次ケプストラム係数、Δ対数パワー及び１６次Δケプストラム係数を含む３４次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ３を介して時間正規化部４に入力される。時間正規化部４は、ターゲット話者の発声音声のスペクトルと、上記音声情報メモリ１０－１乃至１０－Mに予め記

憶された複数M人の話者のスペクトルとの時間整合を、距離尺度としてケプストラム距離を用いてＤＴＷ（Dynamic time warping）法により実行する。すなわち、ターゲット話者の発声音声の例えば単語又は文の時間長さは人及び時々により変化するので、当該ターゲット話者の発生音声のスペクトルのデータを、その単語又は文と同一の単語又は文に関する複数M人の話者の音声スペクトルの時間長さと同一となるように時間整合処理（時間正規化処理）を実行し、処理後のターゲット話者の音声スペクトルのデータは補間比率決定部５に出力される。

【００１９】音声情報メモリ１０－１乃至１０－Mから読み出される複数M人の音声スペクトルのケプストラム係数データは補間比率決定部５に出力されるとともに、乗算器１１－１乃至１１－Mに出力される。乗算器１１－１乃至１１－Mはそれぞれ、入力された各人の音声スペクトルのケプストラム係数データと、補間比率決定部５から出力される補間比率a_1，a_2，…，a_Mとを乗算して加算器１２に出力し、加算器１２は入力されるデータを加算して、加算結果のデータをパラメータ時系列生成部６に出力する。すなわち、複数M個の乗算器１１－１乃至１１－Mと加算器１２とによって音声スペクトルの内挿処理が実行される。

【００２０】補間比率決定部５は、ターゲット話者への声質適応を行う場合、すなわち音声合成後のターゲット話者の発声音声に近似させる場合、ターゲット話者のスペクトルとスペクトルの内挿により生成したスペクトルとの間の距離が最小になるように内挿比率を決定する。具体的には、最適な補間比率a_1，a_2，…，a_Mを次の数１で示す関数Ｆの関数値を最小２乗化法により演算して決定する。すなわち、ターゲット話者の発声音声のケプストラム係数値と、それに時間的に対応する予め格納された複数M人の音声スペクトルのケプストラム係数値との差の二乗が最小になるように、補間比率a_1，a_2，…，a_Mを求める。

【００２１】

【数１】

$$F(a_1, a_2, \cdots, a_M) = \sum_{i\ j} (Y_{ij} - y_{ij})^2$$

ここで、

【数２】

$$y_{ij} = a_1 \cdot x_{1ij} + a_2 \cdot x_{2ij} + \cdots + a_M \cdot x_{Mij}$$

ただし、

【数３】

$$\sum_k a_k = 1$$

である。

【００２２】ここで、Y_{ij}とy_{ij}はそれぞれ、ターゲット話者と内挿により得られるスペクトルの第ｉフレームのｊ次ケプストラム係数を表わす。x_{kij}は予め音声情報メモリ１０－１乃至１０－Mに格納されたｋ番目の話者の第ｉフレームのｊ次ケプストラム係数を表わす。なお、音声合成後のターゲット話者の発声音声に近似させず、ターゲット話者とは異なる声質の音声を合成する場合は、補間比率を適宜変更する。この場合、補間比率は、操作者がキーボード（図示せず。）を用いて補間比率決定部５に入力するように構成してもよい。

【００２３】上記補間比率決定部５の処理の後に、パラメータ時系列生成部６は、加算器１２から逐次出力される１６次のケプストラム係数の時系列を取りまとめて、内蔵のバッファメモリに格納した後、そのデータを係数変換部７に出力する。係数変換部７は、入力された１６次のケプストラム係数の時系列データに基づいて、そのデータを、公知の方法により、音声合成のためのピッチ、有声／無声切り換え、振幅及びフィルタ係数のデータに変換して、それぞれパルス発生器２１とスイッチＳＷと振幅変更型増幅器２３とフィルタ２４とに出力する。

【００２４】音声合成部２０は、パルス発生器２１と雑音発生器２２とスイッチＳＷと振幅変更型増幅器２３とフィルタ２４とから構成される。パルス発生器２１は、有声音の励振音源であって各ピッチ周期の開始時点で単位大きさのインパルスを発生して、スイッチＳＷを介して振幅変更型増幅器２３に出力する。一方、雑音発生器２２は、無声音の励振音源であって、無相関でかつ一様分布を有する標準偏差１と平均値０のランダム雑音を発生して、スイッチＳＷを介して振幅変更型増幅器２３に出力する。従って、スイッチＳＷは有声音を発生するときパルス発生器２１側に切り換える一方、無声音を発生するときは雑音発生器２２側に切り換られる。さらに、振幅変更型増幅器２３は、入力される振幅情報に基づいて入力される信号の振幅を変更しかつ増幅してフィルタ２４に出力する。そして、フィルタ２４は、その伝達関数に対応するフィルタ係数を入力されるフィルタ係数に設定し、入力された信号を当該設定されたフィルタ係数でろ波した後、スピーカ３０を介して出力する。

【００２５】以上実施例において、音声合成後のターゲット話者の発声音声に近似させる場合、このスピーカ３０からは、ターゲット話者の発声音声に近似した音声信号が出力される一方、音声合成後のターゲット話者の発声音声に近似させず、ターゲット話者とは異なる声質の音声を合成する場合は、設定された補間比率に対応して声質制御された音声信号を出力させることができる。なお、後者の場合においては、ターゲット話者の発生音声に基づかず、補間比率決定部５以降の回路のみで構成してもよい。また、音声情報メモリ１０－１乃至１０－Mに格納される音声スペクトルのデータは予め時間正規化処理を実行されていることが好ましい。

【００２６】以上の実施例において、スペクトル・パラメータとして、ケプストラム係数を用いているが、本発明はこれに限らず、ＰＡＲＣＯＲパラメータk_iから誘

7

導できる等価パラメータ集合を示す次の数４の対数面積比パラメータg_iを用いてもよい。この場合、補間比率決定部５は、音声合成後のターゲット話者の発声音声に近似させるとき、好ましくは、数１を用いて演算した補＊

$$g_i = \log [(1-k_i) / (1+k_i)], \quad 1 \leq i \leq p$$

【００２７】本発明者は、本実施例の装置を用いてシミュレーションを実行して、２話者ので間で補間が良好に行えるかどうかを調べた。音質に関しては、ケプストラム係数と対数面積比のどちらのパラメータを用いた場合も比較的良好に補間して内挿することができ、補間比率を変化することにより声質が一方の話者から他方の話者に安定に徐々に変化していくことを確認した。

【００２８】

【発明の効果】以上詳述したように本発明によれば、複数の話者の音声スペクトルのデータを予め記憶する記憶手段と、上記記憶手段から複数の話者の音声スペクトルのデータを読み出し、所定の補間比率を用いて上記複数の話者の音声スペクトルのデータに対して内挿処理を実行して所定の声質を有する音声スペクトルのデータを出力する処理手段と、上記処理手段から出力される音声スペクトルのデータに基づいて音声を合成して出力する音声合成手段とを備えたので、従来例に比較してより小さい記憶容量を有する記憶装置を用い、より簡単な構成の回路を用いて、より自由度が高い声質の制御が可能になり、より多様な声質を有する音声を合成することができる。

【００２９】さらに、ターゲット話者の音声に基づいて音声スペクトルのデータを抽出する特徴抽出手段と、上記特徴抽出手段によって抽出された音声スペクトルのデータが上記処理手段から出力される音声スペクトルのデータに近似するように上記補間比率を演算して上記処理

8

＊間比率を初期値として、さらに、非線形降下法を用いてケプストラム距離の低減を行う。

【数４】

手段に設定する演算手段とを備えたので、ターゲット話者が発声した少量の音声を入力として、声質変換のターゲットとする話者の声質を近似する音声を発生することができるという特有の利点がある。

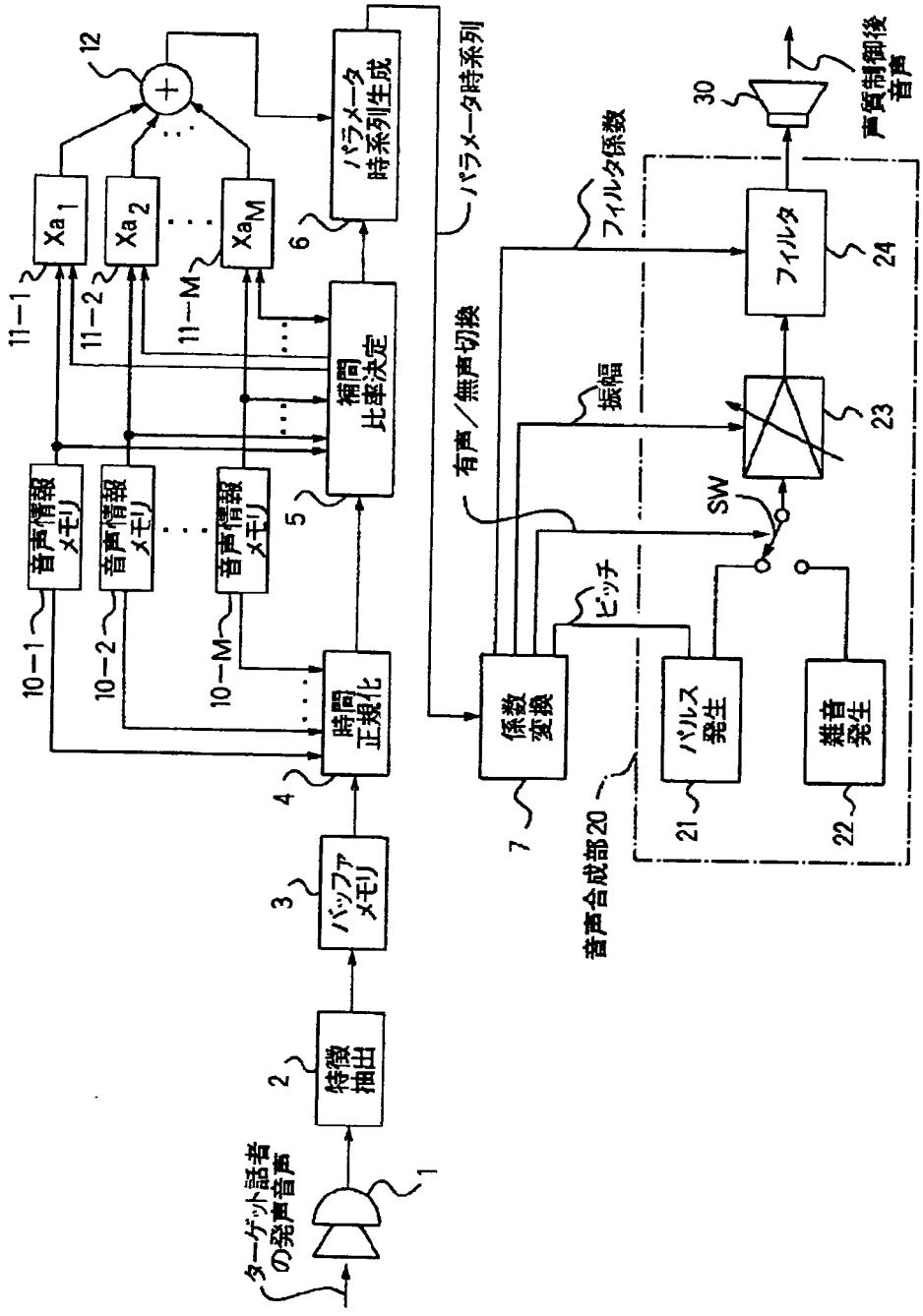
【図面の簡単な説明】

【図１】　本発明に係る一実施例である声質制御型音声合成装置のブロック図である。

【符号の説明】

１…マイクロホン、
２…特徴抽出部、
３…バッファメモリ、
４…時間正規化部、
５…補間比率決定部、
６…パラメータ時系列生成部、
７…係数変換部、
１０－１乃至１０－Ｍ…音声情報メモリ、
１１－１乃至１１－Ｍ…乗算器、
１２…加算器、
２０…音声合成部、
２１…パルス発生器、
２２…雑音発生器、
２３…利得変更型増幅器、
２４…フィルタ、
３０…スピーカ、
ＳＷ…スイッチ。

【図1】


ターゲット話者の発声音声
1
特徴抽出
2
バッファメモリ
3
時間正規化
4
補間比率決定
5
パラメータ時系列生成
6
パラメータ時系列
有声／無声切換
係数変換
7
フィルタ係数
ピッチ
パルス発生
21
雑音発生
22
音声合成部20
振幅
SW
フィルタ
23
24
30
声質制御後音声
音声情報メモリ
10－1
10－2
10－M
11－1
11－2
11－M
Xa1
Xa2
XaM
12